Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: A new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407–4414.
Wachira, F.N., R. Waugh, C.A. Hackett, and W. Powell. 1995. Detection of genetic diversity in tea (*Camellia sinensis*) using RAPD markers. Genome 38:201–210.
Wight, W. 1962. Tea classification revised. Curr. Sci. 31:298–299.
Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531–6535.
Xu, R.Q., N. Tomooka, and D.A. Vaughan. 2000. AFLP markers for characterizing the Azuki bean complex. Crop Sci. 40:808–815.
Yeh, F.C., R.C. Young, B. Timothy, T.B.J. Boyle, Z.H. Ye, and J.X. Mao. 1997. POPGENE, the user-friendly shareware for population genetics analysis. Molecular Biology and Biotechnology Center, University of Alberta, Canada. (http://www.ualberta.ca/~fyeh; verified September 27, 2001).
Zhang, Q., M.A. Saghai Maroof, T.Y. Lu, and B.Z. Shen. 1992. Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. Theor. Appl. Genet. 83:495–499.

Evaluation of Genetic Diversity in Rice Subspecies Using Microsatellite Markers

Junjian Ni, Peter M. Colowit, and David J. Mackill*

ABSTRACT

Molecular markers are useful tools for evaluating genetic diversity and determining cultivar identity. The purpose of this study was to evaluate the genetic diversity within a diverse collection of rice (*Oryza sativa* L.) accessions, and to determine differences in the patterns of diversity within the two rice subspecies *indica* and *japonica*. Thirty-eight rice cultivars of particular interest to U.S. breeding programs and two wild species accessions (*O. rufipogon* Griffithand *O. nivara* Sharma et Shastry) were evaluated by means of 111 microsatellite markers distributed over the whole rice genome. A total of 753 alleles were detected, and the number of alleles per marker ranged from 1 to 17, with an average of 6.8. A positive correlation was found between the number of alleles per locus and the maximum number of repeats within a microsatellite marker. Compared to *indica* cultivars, the *japonica* group showed significantly higher genetic diversity on chromosomes 6 and 7, and considerably lower diversity on chromosome 2. All rice cultivars and lines could be uniquely distinguished, and the resulting groups corresponded exactly to the *indica* and *japonica* subspecies, with *japonica* divided into temperate and tropical types. With stepwise discrimination, two subsets of approximately 30 markers were identified that produced genetic distance matrices and dendrograms that were the same as those produced by means of all 111 markers. The results suggested that a relatively small number of microsatellite markers could be used for the estimation of genetic diversity and the identification of rice cultivars.

Rice has one of the largest ex situ germplasm collections in the world (Jackson and Juggan, 1993). This accessible collection of diverse cultivated and wild rice germplasm has made great contributions to rice breeding. The development of isozyme and, later, DNA marker technology has provided an efficient tool to facilitate plant genetic resource conservation and management. In rice, molecular markers have been used to identify accessions (Olufowote et al., 1997; Virk et al., 1995), to determine the genetic structure and pattern of diversity for cultivars of interest (Akagi et al., 1997; Mackill, 1995; Yang et al., 1994; Zhang et al., 1992), and to optimize the assembly of core collections (Schoen and Brown, 1995). Compared to morphological analysis, molecular markers can reveal differences among accessions at the DNA level and thus provide a more direct, reliable, and efficient tool for germplasm conservation and management.

Several types of molecular markers are available for evaluating the extent of genetic variation in rice. These include restriction fragment length polymorphism (RFLP) (Botstein et al., 1980), random amplified polymorphic DNA (RAPD) (Welsh and McClelland, 1990; Williams et al., 1990), amplified fragment length polymorphism (AFLP) (Vos et al., 1995), and microsatellite or simple sequence repeat (SSR) (Tautz, 1989). Of these, RFLP and microsatellites are codominant markers and their map positions on the rice genome are well known, while RAPD and AFLP markers involve the use of random, largely dominant markers. Microsatellites are PCR-based markers that are both technically efficient and cost-effective to use and are available for rice (Chen et al., 1997; Temnykh et al., 2000). Compared with RFLPs, microsatellite markers detect a significantly higher degree of polymorphism in rice (Wu and Tanksley, 1993; Yang et al., 1994), and are especially suitable for evaluating genetic diversity among closely related rice cultivars (Akagi et al., 1997).

The cultivated Asian rice species, *O. sativa*, is composed of two subspecies, indica and *japonica* (Oka, 1988). *Indica* is the predominant tropical subspecies. The *japonica* subspecies, consisting of temperate and tropical types, is widely grown in East Asia, North and South America, Australia, Mediterranean North Africa, and Europe, and accounts for about 20% of world rice production (Mackill, 1995). The genetic diversity of *japonica* rice is thought to be lower than for *indica* rice (Glaszmann, 1987; Zhang et al., 1992). The use of wide crosses between different subspecies often results in sterility problems in the hybrids and their progenies, disruption of favorable linkage blocks and gene combinations, and linkage drag (Ikehashi and Araki, 1986). The reduced recombination and distorted segregation resulting from wide hybridization may cause difficulties in selection for desired recombinants during the breeding process (Pham and Bougerol, 1993). From the viewpoint of rice breeders, it is preferable to identify and

J. Ni, Dep. of Agronomy & Range Science, Univ. of California, Davis, CA 95616-8515 USA; P.M. Colowit, USDA-ARS, Dep. of Agronomy & Range Science, Univ. of California, Davis, CA 95616-8515 USA; and D.J. Mackill, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines. Received 3 Nov. 2000. *Corresponding author (d.mackill@cgiar.org).

Published in Crop Sci. 42:601–607 (2002).

Abbreviations: AFLP, amplified fragment length polymorphism; MAS, marker assisted selection; RAPD, random amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; simple sequence repeat, SSR.

Table 1. List of rice accessions used to study microsatellite marker variation.

Name	Origin	Subspecies–group
Black Gora	India	*Indica*
IR40931	Philippines	*Indica*
IR50R	IRRI	*Indica*
IR36	IRRI	*Indica*
Hunan Late Indica 2	China	*Indica*
N-22	India	*Indica*
GIZA178	Africa	*Indica*
GZ5121-5-2-1	Africa	*Indica*
GZ5470-14-1-2	Africa	*Indica*
Teqing	China	*Indica*
L-202	USA-Ca	*Japonica*-tropical
L-203	USA-Ca	*Japonica*-tropical
Labelle	USA-So	*Japonica*-tropical
87Y550	USA-Ca	*Japonica*-tropical
Lemont	USA-So	*Japonica*-tropical
Moroberekan	Africa	*Japonica*-tropical
Katy	USA-So	*Japonica*-tropical
Newbonnet	USA-So	*Japonica*-tropical
M-103	USA-Ca	*Japonica*-temperate
M-201	USA-Ca	*Japonica*-temperate
M-202	USA-Ca	*Japonica*-temperate
M-203	USA-Ca	*Japonica*-temperate
M-204	USA-Ca	*Japonica*-temperate
M-401	USA-Ca	*Japonica*-temperate
Italica Livorno	Italy	*Japonica*-temperate
WC 1403	Korea	*Japonica*-temperate
Akitakomachi	Japan	*Japonica*-temperate
Koshihikari	Japan	*Japonica*-temperate
Yukihikari	Japan	*Japonica*-temperate
Arborio	Italy	*Japonica*-temperate
Calmochi 102	USA-CA	*Japonica*-temperate
Calrose	USA-CA	*Japonica*-temperate
Taipei 309	China	*Japonica*-temperate
Uz Ros 269	Russia	*Japonica*-temperate
Daegwanbyeo	Korea	*Japonica*-temperate
S-201	USA-CA	*Japonica*-temperate
Hirayana	Japan	*Japonica*-temperate
GIZA176	Africa	*Japonica*-temperate
O. rufipogon		Wild relative
O. nivara		Wild relative

use donors of important traits from within the same subspecies or cultivar group. For the application of marker assisted selection (MAS) within a subspecies, it is important to obtain information on the genetic diversity within a rice subspecies over different genome regions. The excellent attributes of SSR markers and the availability of over 300 markers in rice (Temnykh et al., 2000) make it possible to obtain this information.

Our objectives were to use microsatellite markers to evaluate the genetic variation within a diverse collection of rice accessions, to determine differences in the patterns of diversity within two rice subspecies, to distinguish different accessions, and to reveal the genetic relationships among them.

MATERIALS AND METHODS

Accessions used in the present study included 10 *indica* cultivars, eight tropical *japonica* cultivars, 20 temperate *japonica* cultivars, and *O. rufipogon* and *O. nivara*, two close relatives of *O. sativa* (Table 1). Accessions were obtained from the Rice Germplasm Center at the International Rice Research Institute (IRRI), Philippines, the Rice Experiment Station, Biggs, CA, and the local collection at Davis, CA.

DNA Extraction and SSR Analysis

DNA extraction from rice leaves was as described in Redoña and Mackill (1998). One hundred-eleven rice microsatellite primer pairs were chosen to represent the entire rice genome at about 15- to 20-centimorgan intervals on the basis of the published rice microsatellite framework map (Temnykh et al., 2000) (Table 2). The original sources and motifs for these markers can be found in Temnykh et al. (2000) and in the RiceGenes database (http://ars-genome.cornell.edu/rice; verified October 2, 2001). Polymerase chain reaction (PCR) analysis followed procedures recommended by the manufacturer (ABI Prism 377 GeneScan Chemistry Guide, PE Biosystems, Foster City, CA) with minor modifications. It was performed in 15 μL of a mixture containing 50 ng DNA, 330 n*M* of each primer, 250 μ*M* of each dNTP, and 0.6 U Taq DNA polymerase in reaction buffer [20 m*M* TRIS pH 8.0, 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 50% (v/v) glycerol]. Fluorescent d CTPs (330 n*M*) labeled with a rhodamine dye (R110, R6G) was incorporated into PCR products to enable detection of the fragments in the ABI377 automated sequencing system (Perkin-Elmer). The PCR was run as follows: (i) an initial denaturation step of 3 min at 94°C, (ii) 35 cycles of 1 min at 94°C, 2 min at 55°C, 1.5 min at 72°C, and (iii) a final extension step for 5 min at 72°C. For some specific cDNA derived microsatellites, two different annealing temperatures, 61 and 67°C, were employed as described by Temnykh et al. (2000). PCR products were analyzed on a sequencing gel (5% LongRanger, 1× TBE buffer, 6 *M* urea) in an automated ABI377 sequencing apparatus (Perkin-Elmer). Fragment lengths were estimated using internal size standards by GeneScan Analysis Software.

Table 2. List of microsatellite markers used in the study. Marker designations are from Chen et al. (1997) and Temnykh et al. (2000).

Chromosome	Microsatellite marker
1	RM5, RM9, RM81A, RM84, RM212, RM220, RM237, RM238A, RM243, RM246, RM259, RM 1, RM 23, RM265, RM315, RM302
2	RM207, RM211, RM233A, RM240, RM250, RM262, RM263, RM6, RM213, RM341, RM266
3	RM232, RM251, RM130, RM22, RM135, RM156, RM55, RM338, RM60
4	RM261, RM241, RM142, RM127, RM177, RM317, RM348, RM307, RM185
5	RM153, RM31, RM169, RM173, RM178, RM13, RM291, RM274, RM334
6	RM30, RM3, RM204, RM253, RM136, RM170, RM340, RM345, RM193
7	RM82, RM234, RM214, RM172, RM51, RM10, RM180
8	RM152, RM137, RM52, RM126, RM223, RM230, RM337, RM264, RM256
9	RM215, RM105, RM219, RM160, RM245, RM296, RM278
10	RM228, RM258, RM244, RM239, RM171, RM222, RM271, RM311
11	RM 21, RM167, RM202, RM209, RM224, RM229, RM254, RM332, RM287
12	RM247, RM235, RM155, RM20, RM17, RM313, RM309, RM270

Data Analysis

The number of repeats for each allele was determined by comparing the size of the PCR product with that of IR36 whose repeat number was characterized by Temnykh et al. (2000). Estimated repeat number was used in the following analysis. The number of alleles per locus was based on an evaluation of the 38 *O*. sativa cultivars and didn't include the two wild species. The term polymorphism information content (PIC) was originally introduced into human genetics by Botstein et al. (1980). It refers to the value of a marker for detecting polymorphism within a population, depending on

Table 3. Mean values for allele number and PIC of different classes of microsatellite markers within rice and by subspecies. Both polymorphic and monomorphic microsatellites were included in the calculations.

Class of microsatellite markers	Number of markers	Number of alleles			PIC value		
		O. sativa	*Indica*	*Japonica*	*O. sativa*	*Indica*	*Japonica*
Genomic libraries	87	7.41a†	4.62a	4.64a	0.650a	0.651a	0.501a
cDNA-derived	24	4.50b	3.38b	2.88b	0.514b	0.456b	0.335b
Poly (GA)n type	70	7.59a	4.66a	4.77a	0.657a	0.657a	0.512a
Other 2-bp	7	6.57ab	4.43a	4.14ab	0.652ab	0.663ab	0.497a
3-bp	27	5.22b	3.81a	3.37b	0.546b	0.504b	0.375a
4-bp	2	3.50b	3.50a	1.50b	0.384b	0.500b	0.250a
Complex	5	5.60ab	3.20a	3.20b	0.554a	0.460b	0.329a
All	111	6.78	4.35	4.26	0.621	0.609	0.465

† Means in a column followed by different letters are statistically different at $P \pm 0.05$.

the number of detectable alleles and the distribution of their frequency. In present study, PIC value of a marker was calculated according to a simplified version after Anderson et al. (1993):

$$PIC_i = 1 - \sum_{j=1}^{n} P_{ij}^2$$

where P_{ij} is the frequency of the jth allele for the ith marker, and summed over n alleles. The PROC REG of SAS was employed to study the relationship between the maximum number of simple repeats and number of alleles or PIC value.

Average number of alleles, average PIC value, and average genetic distance were computed on the basis of different rice subspecies, chromosomes and microsatellite classes, and the means were compared by the GLM (general linear models) procedure of SAS. The Tukey test was used to compare the means of different classes.

The presence (1) and absence (0) of alleles for each microsatellite marker were recorded for all accessions and then converted to a genetic distance matrix. Genetic distances (Nei and Li, 1979) between two entries were computed as

$$GD = 1 - [2N/(N_i + N_j)]$$

where N is the number of shared bands and N_i and N_j are the total number of bands for entries i and j. A cluster diagram was constructed based on these distances by the UPGMA (average linkage) method using PROC CLUSTER of SAS (SAS Institute Inc., 1989). The centroid, median and single linkage clustering methods of PROC CLUSTER were also used to observe consistency of the clustering. Groups determined from the cluster analysis were used for canonical discriminant analysis using the program PROC CANDISC of SAS. Squared Mahalanobis distances between class means were computed and the first two canonical variables were plotted for all accessions. The program PROC STEPDISC of SAS was used to choose the subsets of microsatellite markers that best represented the information in the total dataset. Pearson correlation coefficients (PROC REG of SAS) were used to evaluate the relationship between genetic distances calculated on the basis of sublets and the whole dataset.

RESULTS AND DISCUSSION

SSR Polymorphism in the Entire Sample

The 111 SSR markers revealed 753 alleles in the 38 cultivars. The number of alleles per locus varied widely among these markers, ranging from 1 (RM238A and RM193) to 17 (RM204) with an average of 6.8. These numbers are, on a per locus basis, much larger than those reported from previous studies using other types of markers such as isozymes (Glaszmann, 1987; Second, 1982) and RFLPs (Wang and Tanksley, 1989; Zhang et al., 1992). As a measure of the informativeness of microsatellites, the average PIC value was 0.62 with the range of 0.10 (RM60) to 0.91 (RM204).

The microsatellite markers derived from the genomic library showed significantly higher genetic diversity than those derived from GenBank sequences (Table 3). Comparing microsatellite markers with the different repeat motifs, those with GA repeats had the greatest number of alleles and highest PIC values, while the 3- and 4-bp motif markers had the lowest number of alleles and PIC values (Table 3). These results were consistent with those reported by Cho et al. (2000).

There was a relationship between the number of alleles detected at a locus and the maximum number of simple repeats within the targeted microsatellite DNA ($r = 0.72$, $P < 0.001$). Thus, the larger the maximum repeat number in the microsatellite DNA, the larger the number of alleles detected. A significant correlation between PIC value and the maximum number of single repeats per microsatellite marker was also detected ($r = 0.69$, $P < 0.001$).

Most microsatellite primers amplified PCR products in *O. rufipogon* (89%) and *O. nivara* (88%). In the case of null alleles in these species, PCR amplifications were repeated to exclude failed PCR reaction as the cause of the null allele. The null alleles can arise from point mutation (s) in one or both of the primer sites. The low percentage of null alleles in wild relatives of rice implied that most SSR primers developed for rice could be employed with *O. rufipogon* and *O. nivara*, the close relative of *O. sativa*. For about one fifth of primers capable of being amplified in wild relatives (20 of 99 in *O. rufipogon* and 22 of 98 in *O. nivara*), it was also noted that the band sizes of PCR products of wild relatives were outside the range of those amplified from *O. sativa*, suggesting that those alleles identified in wild relatives might be unique and different from those detected in *O. sativa*.

Comparison of Polymorphism between Rice Subspecies and Chromosomes

Although there were fewer *indica* (10) than *japonica* (28) cultivars included in the present study, the average number of alleles observed was similar in *indica* (4.4) to that in *japonica* (4.3). The *indica* group had significantly greater average PIC value (0.609) than the *japonica*

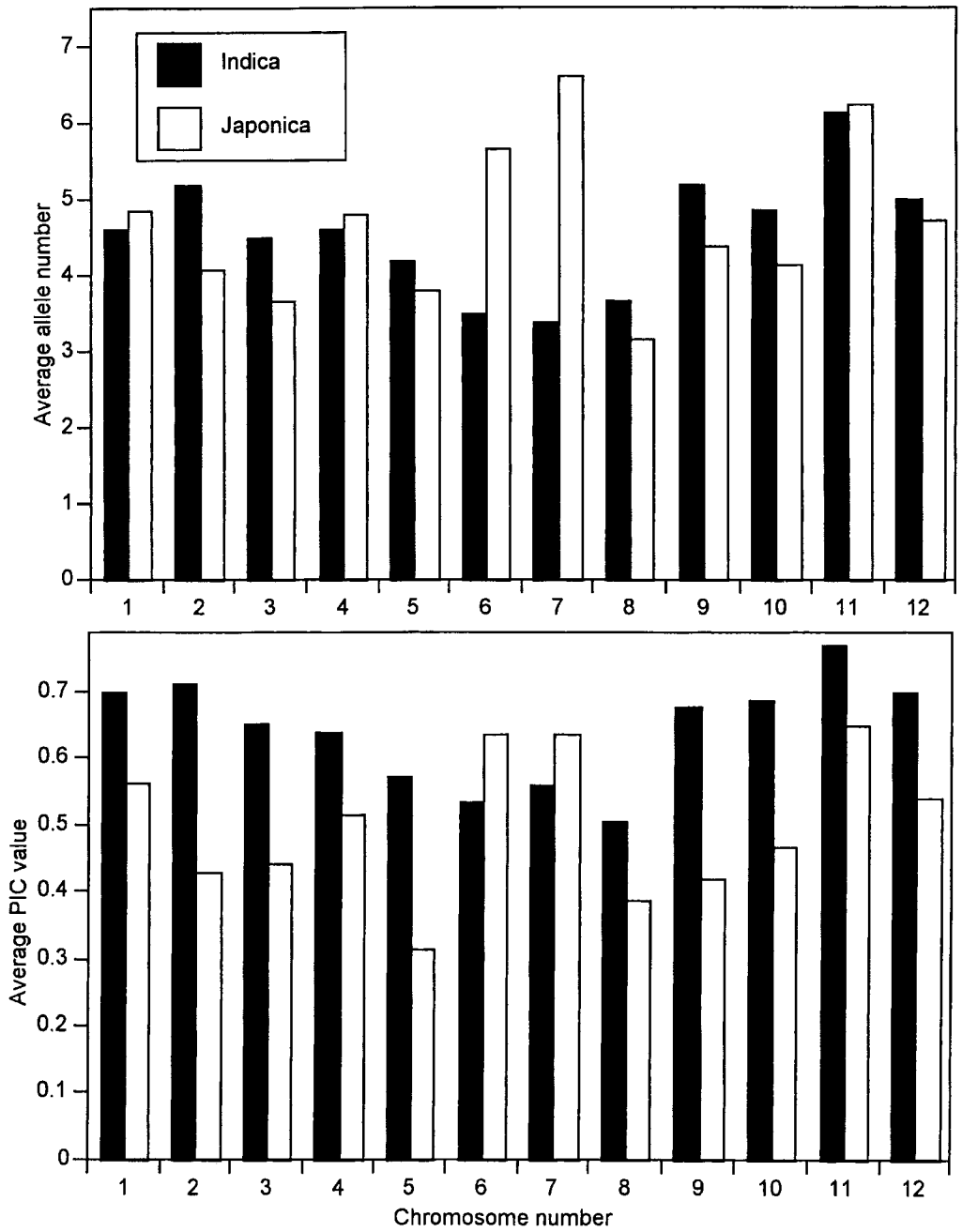

Fig. 1. Comparison of genetic diversity within the *indica* and *japonica* groups based on microsatellite markers on different rice chromosomes. A) Average number of alleles. B) Average of PIC value.

group (0.465) (Table 3). The means of genetic distances between cultivar pairs within the subspecies were also used to evaluate the genetic diversity of different subspecies. The average genetic distance for the *indica* group (0.675) was significantly greater than that for the *japonica* group (0.484) ($F = 140.0$, $P \leq 0.001$). The conclusion that *indica* rice has a higher level of genetic variation than *japonica* rice is in accordance with previous reports (Mackill, 1995; Yang et al., 1994; Zhang et al., 1992). However, it should be noted that several closely related *japonica* cultivars were included.

There has been extensive interest in characterizing the genetic differences between *indica* and *japonica* cultivars, and several studies have been performed to characterize the extent of genetic diversity and differentiation of these two rice groups (Oka, 1988; Second, 1982; Yang et al., 1994; Zhang et al., 1992). Some marker alleles appeared to be diagnostic for rice subspecies. For eight markers (RM240, RM262, RM130, RM156, RM160, RM245, RM271 and RM235), the *indica* cultivars had alleles that were not found in the *japonica* cultivars. For RM130 and RM156 on chromosome 3, all the *indica* cultivars had the same allele, and it was different from any allele found in the *japonica* cultivars. For RM240, RM262, and RM271, all the japonicas had the same allele, which was not be found in the *indica*

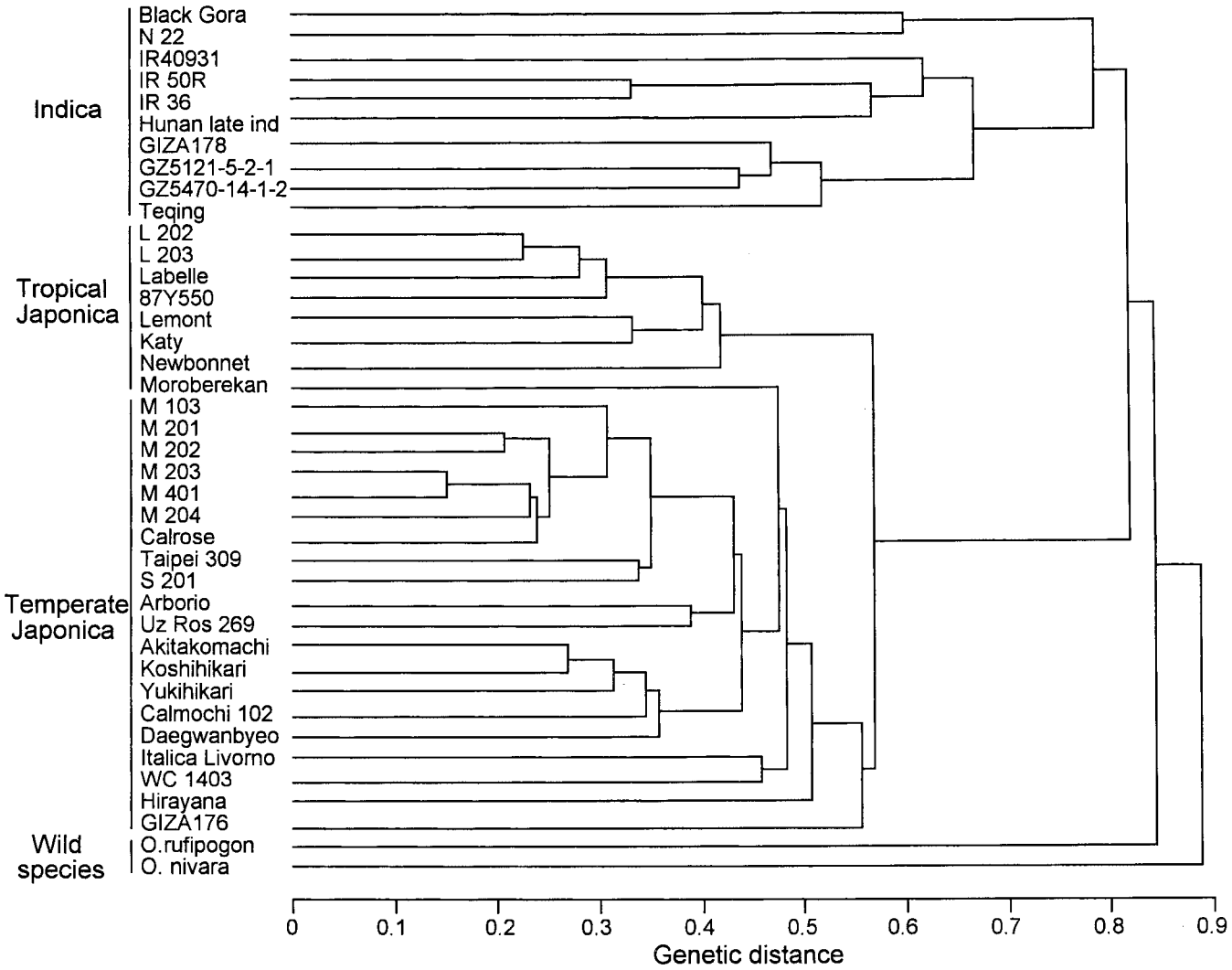

Fig. 2. Cluster diagram based on genetic distance calculated from 111 microsatellite marker alleles of 38 rice cultivars and two wild accessions.

cultivars. In some cases, the allele sizes were sufficiently different to suggest that these alleles could be assessed on agarose gels, which would make them attractive for easily identifying subspecies.

To our knowledge, there is little information available for the difference in genetic diversity of rice subspecies for specific rice chromosomes. In the present study, the number of alleles and PIC value for markers on different chromosomes for *indica* and *japonica* groups were calculated to evaluate the genetic diversity. Both *indica* and *japonica* groups showed a high level genetic variation on chromosome 11 (Fig. 1). The number of alleles on that chromosome was 34% higher than the average based on all 12 chromosomes, and the average PIC values were 20% (*indica*) and 30% (*japonica*) higher. On the other hand, chromosomes 5 and 8 had relatively lower genetic diversity for both groups (Fig. 1). However, diversity of *indica* and *japonica* groups differed on chromosomes 2, 6, and 7. For chromosomes 6 and 7, the *indica* group showed significantly lower genetic diversity than the *japonica* group. In contrast, for chromosome 2, genetic variation of *indica* rice was much higher than that of the *japonica* group (Fig. 1). To exclude complications caused by different microsatellite types, one subset including 87 microsatellite markers derived from the genomic library and another subset including 65 genomic derived $(GA)_n$ microsatellite markers were also used to calculate the means of number of alleles and PIC value on different chromosomes for the two subspecies. Compared with the results based on the whole dataset, similar conclusions were drawn from those two subsets (data not shown).

The information on genetic diversity of rice subspecies for specific genomic regions will be quite useful for rice breeding programs. A major application of this work is to determine the feasibility of mapping genes within the japonica subspecies, and in particular, within the temperate japonica group, to which most California cultivars belong. Some traits such as cooking quality cannot be accurately measured in wide crosses, where the grain quality requirements are completely divergent. The data reported here indicate that it should be possible to obtain adequate polymorphism in crosses between California cultivars and premium quality Japanese cultivars such as Koshihikari to map traits of interest.

For some traits controlled by the specific loci on chro-

Table 4. The two microsatellite marker subsets that fully characterize the rice accessions.

Method	Identified markers
Forward	RM130, RM334, RM173, RM170, RM84, RM256, RM241, RM212, RM13, RM237, RM23, RM258, RM137, RM291, RM52, RM287, RM229, RM213, RM169, RM222, RM60, RM1, RM51, RM127, RM302, RM270, RM9, RM81A, RM232, RM207, RM271
Backward	RM5, RM9, RM81A, RM84, RM212 RM220, RM237, RM243, RM246, RM259, RM1, RM23, RM265, RM315, RM302, RM207, RM211, RM233A, RM240, RM250, RM262, RM263, RM6, RM213, RM341, RM266, RM232, RM251, RM130, RM22, RM135, RM156, RM55, RM338, RM60, RM261, RM241

mosome 6 and 7, it will be easy for a breeder to transfer the desirable trait using MAS by crossing within *japonica* subspecies. However, for the traits controlled by loci on chromosome 2, it might be difficult to obtain markers for use in marker-assisted breeding. Furthermore, the differences in microsatellite diversity may reflect other underlying genetic similarities, implying that one might not find sources of new alleles for traits controlled by genes on this chromosome. Resistance to stem rot (*Sclerotium oryzae* Cattaneo) in rice may be an example of this. Resistance to this disease was not found among *japonica* cultivars, and was introduced from an accession of *O. rufipogon*. One of the important resistance loci was mapped on chromosome 2 (Ni et al., 2001).

Clustering of Rice Cultivars with Microsatellite Markers

All 38 cultivated accessions and the two wild relatives could be easily distinguished even though some accessions were closely related. The UPGMA cluster diagram showed two major clusters corresponding to the *indica* and *japonica* subspecies, with additional sub-clusters within both *indica* and *japonica* clusters (Fig. 2). The same major groups and sub-clusters were observed by the centroid hierarchical, median hierarchical and single linkage clustering methods.

With genetic distance (GD) < 0.56 as the standard for a subcluster, the *japonica* cluster could be divided into two groups. One group was the tropical *japonica* cultivars that included three California long-grain cultivars and four southern U.S. cultivars. The second cluster contained temperate *japonica* cultivars and consisted of two subgroups that remained together in different clustering methods. One subgroup contained typical California medium grain cultivars (M-103, M-201, M-202, M-203, M-401, M-204 and Calrose). The other subgroup mainly contained temperate *japonica* cultivars from East Asia (Akitakomachi, Koshihikari, Yukihikari, and Daegwanbyeo). The remaining temperate *japonica* accessions, however, could not be clustered with these two subgroups. Morobereban, a tropical *japonica* cultivar, was placed in the temperate *japonica* group. This may have occurred because Moroberekan was the only upland cultivar included; all other tropical *japonica* cultivars were U.S. long-grain types. Another explanation is that previous studies have indicated that the boundary between tropical and temperate *japonica* types is not firm (Glaszmann and Arraudeau, 1986; Mackill, 1995). For *indica* subspecies, with the same standard (GD $<$ 0.56), 10 cultivars could be divided into six subclusters (Fig. 2). Considering the high level genetic diversity in the *indica* cultivars, GD < 0.62 was also used for subcluster analysis. In this case, the 10 *indica* accessions could be divided into three subgroups, corresponding to Indian upland cultivars, Egyptian cultivars, and tropical Asian lowland cultivars. On the whole, the clustering results revealed by SSR closely reflected the previously understood relationship among these rice accessions.

The *indica* and the two *japonica* groups determined from the cluster analysis were used to perform the canonical discriminant analysis. The first two canonical variables were plotted to observe the relationship of the three rice cultivar groups (data not shown). By means of 111 microsatellite markers, each group could be clearly distinguished from the others. The squared distances between the *indica* and *japonica* groups were 536 (tropical) and 534 (temperate) compared to 57 between tropical and temperate *japonica*s.

To identify an efficient subset of microsatellite markers, the dataset was subject to STEPDISC analysis. With forward and backward strategies, two subsets of microsatellite markers, consisting of 31 and 37 markers, were identified (Table 4). Average allele numbers were 6.7 and 7.0 for the forward and backward set compared with 6.8 for the whole dataset. PIC values were 0.58 and 0.65 for the forward and backward set compared with 0.62 for the whole dataset. The 31 microsatellites identified from FORWARD STEPDISC analysis were evenly distributed on most of rice chromosomes with the exception of chromosome 11, while the 37 microsatellites from BACKWARD STEPDISC analysis were mainly located on chromosomes 1, 2, 3, and 4. Both subsets included the microsatellites derived from the genomic library and the GenBank sequences. Some markers diagnostic for the rice subspecies, were involved in both subsets (RM130 and RM271 in the FORWARD subset, RM130, RM240, RM262, and RM156 in the BACKWARD subset). These two subsets of markers were used to calculate the genetic distances and cluster analysis. All 38 cultivated accessions and two wild relatives could still be easily distinguished. A significant correlation between the genetic distances calculated on the basis of the total dataset and those based on the subsets was found (for total/forward, $r = 0.983$, $P < 0.001$; for total/backward, $r = 0.971$, $P < 0.001$). The cluster analyses based on these two subsets corresponded exactly to that based on the whole dataset (data not shown). The results suggested that a relatively small number of microsatellite markers could be employed to evaluate the genetic diversity, to identify different accessions, and to reveal the genetic relationship among them.

REFERENCES

Akagi, H., Y. Yokozeki, A. Inagaki, and T. Fujimura. 1997. Highly polymorphic microsatellites of rice consist of AT repeats, and a classification of closely related cultivars with these microsatellite loci. Theor. Appl. Genet. 94:61–67.

Anderson, J.A., G.A. Churchill, J.E. Autrique, S.D. Tanksley, and M.E. Sorrells. 1993. Optimizing parental selection for genetic linkage maps. Genome 36:181–186.

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32:314–331.

Chen, X., S. Temnykh, Y. Xu, Y.G. Cho, and S.R. McCouch. 1997. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor. Appl. Genet. 95:553–567.

Cho, Y.G., T. Ishii, S. Temnykh, X. Chen, L. Lipovich, S.R. McCouch, W.D. Park, N. Ayres, and S. Cartinhour. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100:713–722.

Glaszmann, J.C. 1987. Isozymes and classification of Asian rice varieties. Theor. Appl. Genet. 74:21–30.

Glaszmann, J.C., and M. Arraudeau. 1986. Rice plant type variation: ''Japonica'' - ''Javanica'' relationships. Rice Genet. Newsl. 3:41–43.

Ikehashi, H., and H. Araki. 1986. Genetics of F_1 sterility in rice. p. 119–132. *In* Rice genetics. International Rice Research Institute, Los Baños, The Philippines.

Jackson, M.T., and R. Juggan. 1993. Sharing the diversity of rice to feed the world. Diversity 9:22–25.

Mackill, D.J. 1995. Classifying japonica rice cultivars with RAPD markers. Crop Sci. 35:889–894.

Nei, M., and W.H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. (USA) 76:5269–5273.

Ni, J., P.M. Colowit, J.J. Oster, K. Xu, and D.J. Mackill. 2001. Molecular markers linked to stem rot resistance in rice. Theor. Appl. Genet. 102:511–516.

Oka, H.I. 1988. Origin of cultivated rice. Elsevier, Tokyo.

Olufowote, J.O., Y. Xu, X. Chen, W.D. Park, H.M. Beachell, R.H. Dilday, M. Goto, and S.R. McCouch. 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40:370–378.

Pham, J.L., and B. Bougerol. 1993. Abnormal segregations in crosses between two cultivated rice species. Heredity 70:466–471.

Redoña, E.D., and D.J. Mackill. 1998. Quantitative trait locus analysis for rice panicle and grain characteristics. Theor. Appl. Genet. 96: 957–963.

SAS Institute Inc. 1989. SAS/STAT user's guide, Version 6, Fourth Edition, Volume 1. SAS Institute Inc., Cary, NC.

Schoen, D.J., and A.D.H. Brown. 1995. Maximising genetic diversity in core collections of wild relatives of crop species. p. 55–76. *In* T. Hodgkin et al. (ed.) Core collections of plant genetic resources. John Wiley and Sons, Chichester, UK.

Second, G. 1982. Origin of the genic diversity of cultivated rice (*Oryza* spp.): Study of the polymorphism scored at 40 isozyme loci. Jpn. J. Genet. 57:25–57.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 17:6463–6471.

Temnykh, S., W.D. Park, N. Ayres, S. Cartinhour, N. Hauck, L. Lipovich, Y.G. Cho, T. Ishii, and S.R. McCouch. 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100:697–712.

Virk, P.S., B.V. Ford-Lloyd, M.T. Jackson, and H.J. Newbury. 1995. Use of RAPD for the study of diversity within plant germplasm collections. Heredity 74:170–179.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Fritjers, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new concept for DNA fingerpinting. Nucleic Acids Res. 23:4407–4414.

Wang, Z.Y., and S.D. Tanksley. 1989. Restriction fragment length polymorphism in *Oryza sativa* L. Genome 32:1113–1118.

Welsh, J., and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18:7213–7218.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531–6535.

Wu, K.S., and S.D. Tanksley. 1993. Genetic and physical mapping of telomers and macrosatellites of rice. Plant Mol. Biol. 22:861–872.

Yang, G.P., M.A.S. Maroof, C.G. Xu, Q. Zhang, and R.M. Biyashev. 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol. Gen. Genet. 245:187–194.

Zhang, Q.F., M.A.S. Maroof, T.Y. Lu, and B.Z. Shen. 1992. Genetic diversity and differentiation of indica and japonica rice detected by RFLP analysis. Theor. Appl. Genet. 83:495–499.